A Procedure for Merging Land Cover/Use Data from LANDSAT, Aerial Photography, and Map Sources: Compatibility, Accuracy, and Cost

Prepared for:

1977

ASP - ACSM CONVENTION

WASHINGTON, D.C.

FEBRUARY 27 - MARCH 5

THE AMERICAN SOCIETY OF PHOTOGRAMMETRY

and

THE AMERICAN CONGRESS ON SURVEYING AND MAPPING

Bendix Aerospace Systems Division
3621 South State Road
Ann Arbor, Michigan 48107
Tel (313) 665-7766

Remote Sensing Project
201 UPLA Building
Michigan State University
East Lansing, Michigan 48824
Tel (517) 353-7195

A PROCEDURE FOR MERGING LAND COVER/USE DATA FROM LANDSAT, AERIAL PHOTOGRAPHY, AND MAP SOURCES: COMPATIBILITY, ACCURACY AND COST

W. R. Enslin, S. E. Tilmann, and R. Hill-Rowley

Remote Sensing Project
201 UPLA Building
Michigan State University
East Lansing, Michigan 48824

and

R. H. Rogers
Bendix Aerospace Systems Division
Ann Arbor, Michigan 48107

BIOGRAPHICAL SKETCHES

William R. Enslin is Director of the MSU Remote Sensing Project. He received his M.A. degree in Geography from Eastern Michigan University and is presently working on a Ph.D. at MSU. His research interests and publications center on applying remote sensing technology to land/resource-use problems.

Stephen E. Tilmann holds B.S. and M.S. degrees from Michigan State University and currently is appointed as a Research Specialist with the Remote Sensing Project. He has authored papers concerning geological analysis for land use planning, erosion prediction modelling, engineering characteristics of soil, and ground water recharge.

Richard Hill-Rowley is currently a graduate research assistant with the Remote Sensing Project at Michigan State University. He received his M.A. degree in Geography from the University of Georgia and is currently a Ph.D. candidate in Geography at MSU.

Dr. Robert H. Rogers is a senior staff engineer at Bendix, where he is a Principal Investigator for NASA/ERTS and Skylab/EREP programs. Rogers received his B.S. from Tri-State College, his M.S. from Southern Methodist University, and his Ph.D. in electrical engineering from Michigan State University. He has published many papers in the field of Remote Sensing.

ABSTRACT

Regional planning agencies are currently expressing a need for detailed land cover/use information to effectively meet the requirements of various federal programs. Individual data sources have advantages and limitations in fulfilling this need, both in terms of time/cost and technological capability. A methodology has been developed to merge land cover/use data from LANDSAT, aerial photography and map sources to maximize the effective use of a variety of data sources in the provision of an integrated information system for regional analysis.

A test of the proposed inventory method is currently under way in four central Michigan townships. This test will evaluate the compatibility, accuracy and cost of the integrated method with reference

to inventories developed from a single data source, and determine both the technological feasibility and analytical potential of such a system.

INTRODUCTION

Current and accurate land cover/use information is a basic component of natural resource analyses and the land planning process. In fact, its acquisition is often mandated by new legislation and governmental programs, e.g. Section 208 of the Federal Water Pollution Control Act Amendments of 1972. Recently, many regional planning agencies in Michigan (under Section 208 funding) have acquired land cover/use information through either computer-assisted categorization of LANDSAT data or manual interpretation of aerial photography. Concurrently, there is an increasing demand for interfacing the remotely-sensed inventory data with existing geocoded information storage, analysis, and retrieval systems.

Both photographic and satellite data sources have advantages and limitations with respect to providing all data elements in an accurate cost-effective manner. LANDSAT data processing is a least-cost method of producing general land cover maps and tabular data for large areas. Planning studies, however, often require more detailed land cover/use information at an accuracy level that is difficult to provide consistently over a range of categories through the LANDSAT data extractive process. Agencies in Michigan are also using grid-based information systems which necessitates a raster to grid-cell conversion for LANDSAT data. Manual interpretation of aerial photography is a more expensive and time-consuming process than digital multispectral processing, but it yields a more detailed categorization of land cover/uses that many planning activities require. Many of these categories appear to be unobtainable by LANDSAT.

There is a need, therefore, to establish an appropriate combination of procedures that maximize the effective use of both data sources in the provision of an integrated information system for regional analyses. This paper discusses such a procedure that is currently being investigated and developed at Michigan State University (MSU).

The overall objective of the program described in this report is to provide a grid-based land cover/use data file that is responsive to the specific needs of regional planning agencies by utilizing a cost-effective combination of data capture procedures. This is accomplished through the blending of interpreted results achieved by computer-assisted categorization of LANDSAT data and manual interpretation of aerial photography and maps.

METHODOLOGY

A method has been developed to merge land cover/use data from LANDSAT, aerial photography and map sources into a grid-based geographic information system. The method was developed through expanding and revising an initial strategy documented by Rogers and Tilmann.[1/] The method basically involves:

1. computer-assisted categorization of LANDSAT data to provide certain user-specified land cover categories;

2. manual interpretation of aerial photography to identify other selected land cover/use categories that cannot be obtained from LANDSAT data;

3. identification of special features from aerial photography or map sources;

4. merging of the interpreted data from all the sources into a computer compatible file under a standardized coding structure; and

5. the production of land cover/use maps, thematic maps, and tabular data.

Digital land cover/use files must be structured for immediate use by the planning agency, for the production of map and data products responsive to its program needs. The design of the tape files and the interface with a Natural Resources Information System (NRIS) that manipulates the files are considered jointly by this program.

A NRIS can be configured along several widely differing strategies closely tied into the method of geographic referencing of the inventory data, data analyses, and methods of display. All NRIS, however, have a primary function of providing regional planners with access to a variety of landscape-related data, a means by which to synthesize these data, and a method(s) to display the synthesis results in a meaningful format.

The compatibility of land cover/use data to other data in a NRIS must be defined in at least three ways: the compatibility of the data storage format, the compatiblity of the geographic referencing, and the compatiblity of the types of information stored for the analysis models that are to be used. The program being developed produces a tape file and record structure that carefully considers the format, geographic referencing, and types of information stored for the analysis models that are to be used. The specific NRIS used in this program is the Resource Analysis Program (RAP)2/, although other analytical procedures can be incorporated.

The specific tasks accomplished in producing the merged land cover/use data file and subsequent output products are identified in Figure 1. Note in the flow diagram how the tasks are divided by a vertical dotted line into two parts. Those tasks on the right side of the line are primarily concerned with the interpretation and geocoding of data from photography and maps, and subsequent data analysis and output. Those on the left side are associated with the computer processing of LANDSAT data and the production of the final color-coded maps. Approximate time flow associated with the flow diagram is from top to bottom. The thirteen tasks are briefly described below.

1. The first task provides ground truth data, i.e., aerial photography, maps, and other available data needed to support computer processing of the LANDSAT data on the Bendix Multispectral Data Analysis System (MDAS) or other comparable systems.

2. The ground truth information is used to locate and designate to the computer "training areas" that best typify the land cover categories of interest. The LANDSAT spectral measurements of the training

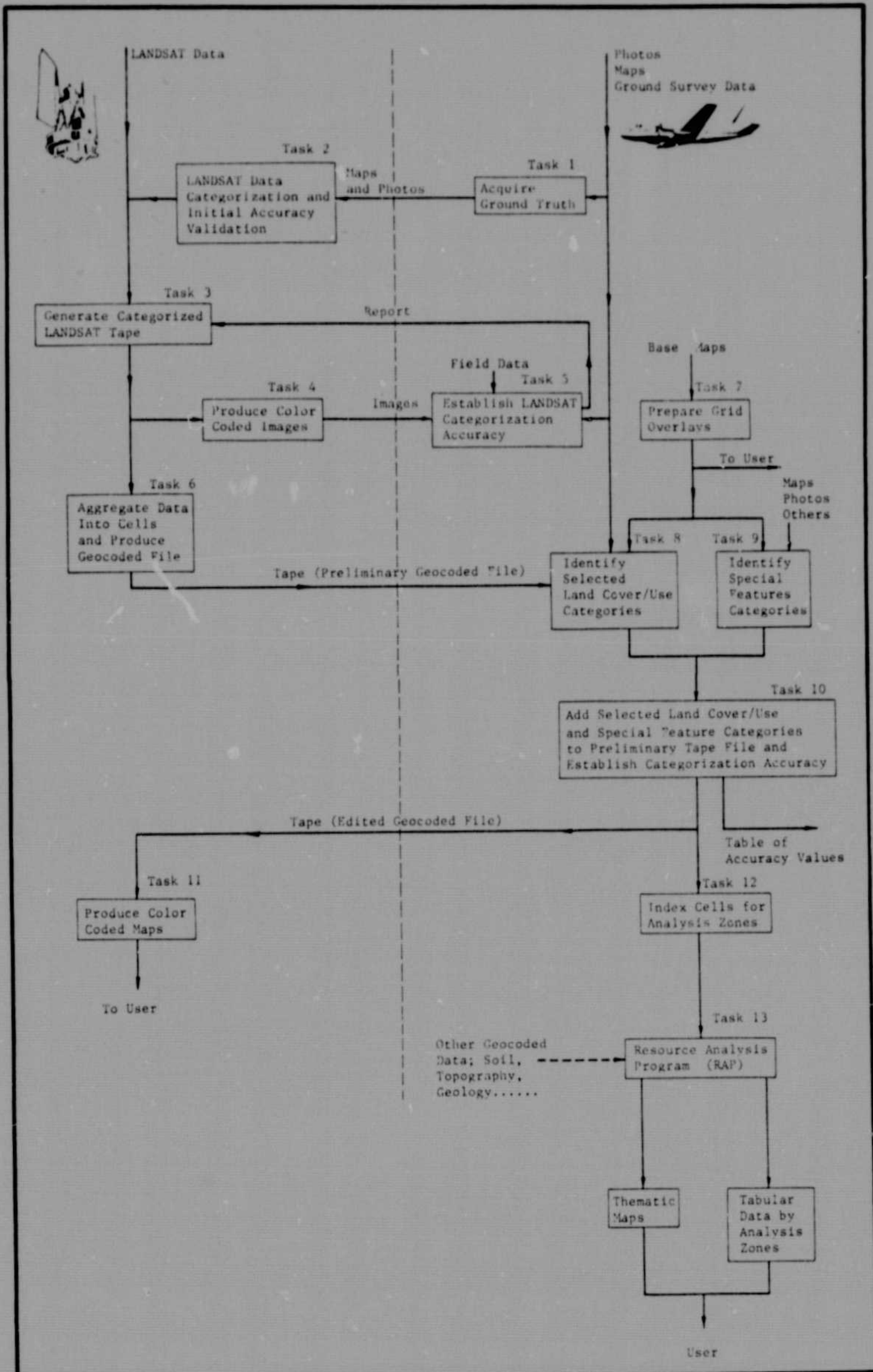


Figure 1. Flow Diagram

areas are then used to generate a set of processing coefficients for each category. Prior to producing categorized data for an entire region, a number of tests are applied to evaluate the computer's ability to perform the desired interpretation, including initial accuracy validation.

3. LANDSAT Computer Compatible Tapes (CCTs) for the planning area are processed on MDAS (using the defined processing coefficients) to produce interpreted land cover data. Throughout the processing, the data are edited and cross-referenced with available ground truth. The selected categories are prepared and documented for each scene with accuracy levels attached to each category. This task results in a "categorized tape". A digital code on this new tape is used to represent the interpreted land-cover category within each picture element (pixel). On the raw data tape, a pixel corresponds to a ground coverage of 57x79 meters or 0.45 hectares (1.1 acres).

4. The LANDSAT categorized tapes are used to produce a color-categorized image of the planning area in which color is used as a code to designate the different land cover categories. The image is produced at the full pixel resolution and shows the region at a user-specified scale (e.g. 1:250,000) and is of suitable quality to verify categorization accuracy and use as an interim product.

5. The color-coded imagery is used in laboratory and field analysis with available photographic coverage and other available ground truth data to confirm LANDSAT categorization accuracy.

6. The categorized LANDSAT data are aggregated into cells. This involves a raster to grid-cell conversion. The initial land cover categories are grouped and the dominant category per cell determined.

7. Transparent grid overlays are produced to fit selected base maps and aerial photographs. These are used to grid-geocode photographic and map data. The grids, with instructive material describing the geocoding procedure, also enables planning personnel to geocode soils, topographic, and geological data so that these data correspond, point by point, with the land cover/use data.

8. The user-specified land cover/use categories that LANDSAT cannot provide are identified through manual interpretation of aerial photography and/or other sources. The grid overlays are then used to geocode the data either by the dominant category within the cell or the category below a cell-centered point.

9. This task uses any previously compiled inventory data (e.g. county drain maps, topographic data, and outdoor recreation maps) or aerial photography and the grid-indexing overlays to interpret and code special features of interest, such as recreational lands, land-fill sites, percent impervious materials, roads, railroads, and cemeteries. Special features occurring within a grid cell are recorded by a unique identifier code. Some cells may contain several special feature codes.

10. The categories derived from the interpretation of photography and maps are added to the geocoded file developed from LANDSAT data. Some LANDSAT land cover codes may be redefined during the merging process. The accuracy of the data file is established through a

sampling procedure which entails random selection of individual cells, a line printer location map of these cells, and ground verification. The output of this task includes the updated geocoded file on nine-track 800-bpi CCTs and a report documenting accuracy.

11. The geo-coded file is used to produce color-coded maps and overlays at a user-specified scale for the region and subareas. On these maps and overlays, color is used to designate the different land cover/use categories. Special feature categories, such as landfill sites, sewage treatment facilities can be placed directly over the color-coded maps and overlays to provide annotations. The details of these composite maps, together with selection of color for categories and design of border annotations and explanation blocks are developed to user specifications.

12. This task codes each cell as to the analysis zone in which it is located (e.g. townships, watershed basins). To accomplish this, the user provides a map showing the desired analysis zones. Boundaries of these zones are digitized and appended to the geocoded tape file by computer techniques. The cell format provides a file structure that can be used immediately to retrieve data by analysis zones.

13. At this point, a digital file containing various land cover/use codes has been prepared. This file is in a standardized format, suitable for use in analytical and mapping programs. Additional resource data may be added to the cell records, such as soil, geological, and topographic data. One example of a program capable of integrating the resource data is the Resource Analysis Program (RAP) developed at MSU. RAP is a user-oriented software system specifically designed to assist regional planning and resource management studies. The graphic capabilities of the program include either line printer or plotter maps for displaying analytical results. In addition, area tabulations can be obtained; for example, the acreage distribution for soil/land cover co-occurrence.

IMPLEMENTATION

Effective implementation of the merging method (outlined in Figure 1) is crucially dependent upon selecting the "best" data source for each user-specified category in terms of accuracy and time/cost trade-offs. A test of an implementation procedure is currently underway that has the following three objectives:

1. determine which categories can be consistently identified through routine LANDSAT data categorization procedures at an adequate level of accuracy (85 to 90 percent);

2. determine appropriate alternative data capture procedures employing aerial photography and maps for those categories that LANDSAT cannot accurately identify;

3. evaluate the accuracy and time/cost associated with operationally applying the integrated inventory procedure with reference to inventories developed from a single data source.

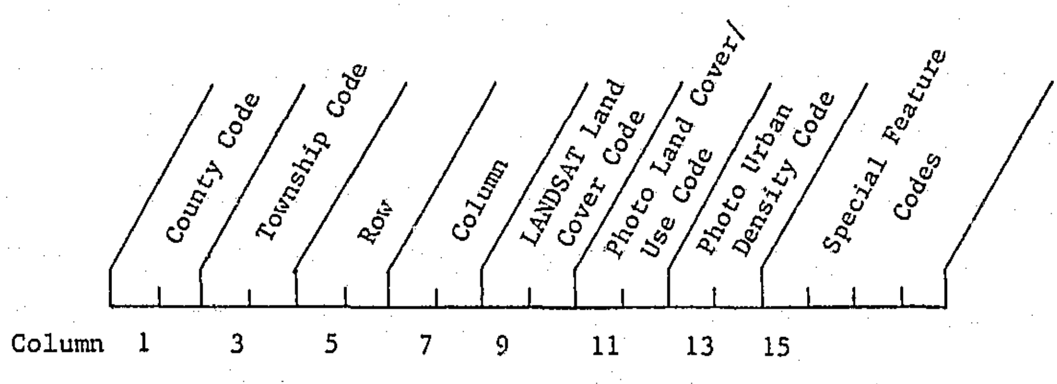
To accomplish objective "1", land cover/use data has been assembled for four townships in central Michigan, representing a diversity of land cover/use conditions. The townships were inventoried through

both LANDSAT data processing and manual interpretation of 1:60,000 and 1:120,000 NASA color infrared photography acquired in 1975.

The land cover/use classification scheme (Table 1) consists of both standard Level I cover categories and "typical" Level II-type categories, as well as selected special features of interest beyond these levels. The categories are defined in the four-level Michigan Land Cover/Use Classification System.[3/] The Michigan System is compatible with the USGS classification system developed for use with remote sensor data.[4/] The photo-derived inventory identified all Level II-type categories. It is assumed that Level II (from photo-interpretation) categories can be aggregated into Level I (from LANDSAT) classes. However, this assertion will ., verified during subsequent analysis. The categorized LANDSAT data is at Level I, except for forest lands which are identified as pine, broadleaf, or mixed forest. The LANDSAT file also contains a measure of urban density (percent impervious materials), i.e., urban lands are categorized as low, medium, or high density. These categories are considered special-interest features.

Both data sets have been geocoded to the dominant category per grid cell. The general structure of a typical cell record is given in Figure 2. Each township constitutes an individual tape file. A standard grid-based geographic referencing system was used with a cell size of 4 hectares (10 acres). Each cell was coded as to dominant land cover (Level I), land cover/use (Level II), and urban density. Additional special feature categories will be added to the file later in the program.

FIGURE 2 STRUCTURE OF TYPICAL CELL RECORD



At this time the registration between the LANDSAT and photointerpreted files are being evaluated. The next step is to identify cells miscategorized by LANDSAT on the basis of the photo inventory. To accomplish this task a cell-level category matrix will be generated from the combined geocoded file. This matrix cross-tabulates the Level I land cover categories identified from LANDSAT data with the categories (Level II-type) derived from photointerpretation.

TABLE 1

LAND COVER/USE CATEGORIES	
1. Urban or Built-up Land	
11	Residential
12	Commercial, Services & Institutional
13	Industrial
14	Transportation, Communication & Utilities
17	Extractive
19	Open & Other
2. Agricultural Land	
21	Cropland, Rotation and Permanent Pasture
221	Tree Fruits
222	Bush-Fruits and Vineyards
23	Confined Feeding Operations
3. Rangeland	
31	Herbaceous Rangeland
32	Shrub Rangeland
4. Forest Land	
41	Broadleaved Forest (generally deciduous)
42	Coniferous Forest
43	Mixed Conifer-Broadleaved Forest
5. Water	
51	Streams & Waterways
52	Lakes
53	Reservoirs
6. Wetland	
61	Forested (wooded) Wetlands
612	Shrub Swamps
62	Non-Forested (non-wooded) Wetlands
7. Barren Land	
74	Barren Transitional Areas
EXAMPLES OF SPECIAL FEATURES	
142	railroads
144	roads
193	outdoor recreation
1464	solid waste disposal sites
1465	sewage treatment facilities
.	percent impervious materials
.	percent vegetative cover
.	others

Analysis of this matrix indicates the initial overall compatibility of the two categorizations--how well Level II-type photo interpretation categories collapse into Level I-type LANDSAT generated categories. More importantly, the analysis allows the pinpointing of cells that, on the basis of photointerpretive information, have been mis-categorized by LANDSAT processing. The pixel components of the identified cells will be examined to define the types of miscategorization. This requires the acquisition of another data set at this resolution level.

The pixel components of the miscategorized cells are extracted from the categorized LANDSAT tape. A corresponding pixel-level inventory is then derived from the aerial photography. This involves reinterpretation of the cells using a .45 hectare minimum type size. A pixel category matrix is then produced which cross-tabulates the categorization by LANDSAT with that by photo interpretation. This matrix will be analyzed along the same lines as the cell-level matrix.

At a number of points in the analysis, redefinition of training sets or changes in decision rules for categorization and aggregation may have to be made. This implies an iteration and re-testing of steps 1 through 6 indicated in the flow diagram (Figure 1). The analysis of the category matrices and subsequent iterations will eventually determine which category can be consistently provided from LANDSAT data processing. The corollary of this is that the remaining categories will have to be identified from aerial photography or existing map sources.

Selecting existing alternative data capture procedures or developing new ones that will provide these categories is the second objective of the program. The specific procedures employed will depend upon the type of land cover/use to be inventoried. After the alternative procedures are defined, the established integrated inventory method will be applied in another area to evaluate associated accuracy and time/cost data with reference to inventories developed from a single data source.

CONCLUDING REMARKS

The methodology described in this paper provides a framework by which users can integrate selected data elements from LANDSAT, photographic, and map sources. The procedures and associated computer software to accomplish integration have been developed, however, effective implementation requires identifying the optimum data source for each land cover/use category. This is a complex issue which has warranted further investigation prior to operationally applying the integrated inventory method. To determine the "best" data capture source and procedure for each category, both LANDSAT and photo categorization of land cover/use in four townships in central Michigan are being analyzed. The objective is to identify a combination of data sources and procedures that provide a more detailed, accurate, and cost-effective land cover/use inventory compared with deriving all the data from a single source.

REFERENCES

(1) Rogers, R.H. and Tilmann, S.E., "A Strategy for Land Use/Cover Data File and Map Preparation." Report BSR 4239, Bendix Aerospace Systems Division, Ann Arbor, MI, May 3, 1976. (Mimeographed.)

(2) Tilmann, S.E., Enslin, W.R., and Hill-Rowley, R. "A Computer Software System for Integration and Analysis of Grid-Based Remote Sensing Data with Other Natural Resources Data". Presented at the 42nd ASP-SCSM Annual Meeting, Washington, D.C., 1977.

(3) Michigan Land Cover/Use Classification System. Michigan Land Use Classification and Referencing Committee. Office of Land Use, Department of Natural Resources. Lansing, Michigan. July 1975.

(4) Anderson, J.R., "A Land-Use Classification System for Use with Remote-Sensor Data", U.S. Geological Survey Circular 671. U.S. Department of Interior, 1972.

ACKNOWLEDGEMENT

This research was supported in part by a National Aeronautics and Space Administration grant, NASA NGL 23-004-083, to Michigan State University, Remote Sensing Project.